DOCUMENT RESUME

ED 140 280 CS 003 525

AUTHOR Potts, George R.
TITLE Drawing Inferences from a Passage of Texts.
INSTITUTION Dartmouth Coll., Hanover, N.H.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C. Office of Research Grants.
BUREAU NO BR-3-0450
PUB DATE 30 Mar 75
GRANT NE-6-00-3-0170
NOTE 32p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Higher Education; *Learning Processes; *Memory; Prose; *Reading Comprehension; *Reading Processes; *Reading Research; Recall (Psychological); Recognition
IDENTIFIERS *Prose Learning

ABSTRACT

The present series of experiments was designed to examine the factors affecting the ability of people to draw inferences from a passage of text. It was found that, using a true-false recognition test, proportion correct was higher and reaction time shorter on inferred information than on information that was actually presented. This was the case for both linear orderings (e.g., "A is better that B. B is better than C.") and set inclusion relations (e.g., "All A are B. All B are C.") This strongly indicates that, when learning a paragraph describing either of these types of ordered information, subjects make and store inferences as an integral part of the reading process. Contrasted with these recognition results, recall of inferences was found to be very poor for both types of relation. Training subjects in such a way as to lead them to expect a recall test did not improve their recall of inferences. This kind of difference between recall and recognition performance is often interpreted as an indication that, although the information in question is stored, it is not easily retrievable. Further research is necessary to determine if this is the case in the above experiments. (Author)

ED140280

FINAL REPORT

Project No. 3-0450

Grant No. NE-6-00-3-0170

Drawing Inferences from a Passage of Text

George R. Potts

Dartmouth College

Hanover, New Hampshire 03755

March 30, 1975

The research reported herein was performed pursuant to a grant with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

U.S. Department of Health, Education and Welfare

National Institute of Education

Office of Research Grants

CS003525

Drawing Inferences from a Passage of Text

George R. Potts

Dartmouth College, Hanover, New Hampshire 03775

Abstract

The present series of experiments was designed to examine the factors affecting the ability of people to draw inferences from a passage of text. It was found that, using a true-false recognition test, proportion correct was higher and reaction time shorter on inferred information than on information that was actually presented. This was the case for both linear orderings (e.g. "A is better than B. B is better than C.") and set inclusion relations (e.g. "All A are B. All B are C."). This strongly indicates that when learning a paragraph describing either of these types of ordered information, subjects make and store inferences as an integral part of the reading process. Contrasted with these recognition results, recall of inferences was found to be very poor for both types of relation. Training subjects in such a way as to lead them to expect a recall test did not serve to improve their recall of inferences. This kind of a difference between recall and recognition performance is often interpreted as an indication that although the information in question is stored, it is not easily retrievable. Further research is necessary to determine if this is the case in the above experiments.

Table of Contents

Drawing Inferences from a Passage of Text

George R. Potts

Dartmouth College, Hanover, New Hampshire 03755

Traditionally, experimental psychologists interested in memory have assumed that learning consists of the passive transfer of certain information from the presentation medium into some internal storage medium. Going beyond the presented information, drawing inferences and coming to conclusions, was viewed as an automatic process which occurred passively as associations were established between various parts of this information. Thus, for example, Hermann Ebbinghaus contended that learning a serial list consisted of learning a chain of pairwise associations between the adjacent items. As this was happening, however, subjects also spontaneously formed more remote associations between non-adjacent items in the list.

It should not be surprising that this approach, taken directly from British Empiricist Philosophy, proved less than useful to educational psychologists who viewed the process of going beyond the text as being at least as important, if not more important, than memorizing the material which was actually presented. Such a theory provided a grossly inadequate framework for anyone interested in examining the processes involved in drawing conclusions from a text, and in discovering what could be done to the material or to the general learning situation to facilitate these processes. In what follows, memory for information which was actually presented will be referred to as <u>reproductive</u> memory; memory for information which must be deduced from the presented information will be referred to as <u>productive</u> memory.

A more useful framework for studying productive memory is provided by the cognitive theories of memory which have just recently become popular among experimental psychologists. These theories have tended to adopt and refine Bartlett's (1932) account of the memory process. According to Bartlett, incoming information is incorporated into "active, organized settings" called schemata. When asked a question about the presented material, subjects must use these active schemata to reconstruct the required information. In Bartlett's words, ". . . the organism would say, if it were able to express itself: 'This and this and this must have occurred, in order that my present state should be what it is' " (p. 202).

Once one assumes that memory is an active process which is at least partially under subjects' control, it becomes reasonable to try to separate the reproductive and productive components of memory and to attempt to examine the various factors which might affect them differentially. Unfortunately, though Bartlett's notion of determination by schemata seems to be a reasonable basis for a cognitive

theory of memory, it is clear that Bartlett's account is not adequately formalized and cannot be made to yield testable predictions. Before we can begin to evaluate the roles of productive and reproductive memory in such a theory, we must examine some of the more recent attempts to clarify the theory and to specify exactly what it is that subjects store when trying to encode meaningful verbal material.

One such attempt was the proposal (e.g. Miller, 1962; Mehler, 1963) that subjects transform each incoming sentence into its linguistic deep structure or sentence kernel (Chomsky, 1957, 1965), and then store that kernel along with a list of the transformations which would be necessary to regenerate the original sentence from the stored kernel. Mehler (1963) referred to this coding strategy as a schema-plus-correction strategy, with the stored kernel corresponding to the schema and the list of transformations comprising a set of corrections to that basic schema. While some researchers have questioned the need for a transformational theory of individual sentence memory (e.g. Martin & Roberts, 1966), a large number of studies have been reported which seem to support this type of linguistic theory (e.g. Mehler, 1963; Mehler & Miller, 1964; Gough, 1965; Savin & Perchonock, 1965; Blumenthal, 1966; Blumenthal & Boakes, 1967; Sachs, 1967; Rohrman, 1968).

It has recently become clear, however, that though linguistic processing of individual sentences plays an important role in the initial stages of comprehension, the final representation of linguistic material does not correspond to a simple catalog of the representations of individual sentences. People's information processing strategies extend far beyond merely altering the form of each incoming sentence and storing that altered form. This has been demonstrated most clearly in a series of studies by Bransford and Franks (1971), Bransford, Barclay, and Franks (1972), and Barclay (1973). During the acquisition phase of these experiments, subjects were presented with a series of sentences, being told to read each for comprehension. After the whole set of acquisition sentences had been presented, subjects were shown another series of sentences, some of which were identical to one of the acquisition sentences and some of which were not. Their task was to indicate whether each test sentence was or was not word-for-word identical to any of the acquisition sentences. It was found that if a test sentence contradicted any information that had been presented during acquisition, subjects were very accurate in recognizing that the sentence had not been presented. As long as the test sentence was not inconsistent with any of the acquisition sentences, however, subjects were unable to make the desired discrimination. Specifically, subjects were unable to distinguish between information which had actually been presented and information which they themselves had deduced from the presented information. Assume, for example, that subjects are first presented with the sentences "The box is to the left of the tree" and "The lamp is on top of the box." When later asked if they had seen the sentence "The lamp is to the left of the tree," subjects will tend to erroneously indicate that they had indeed seen the sentence before. Paris and Carter (1973) have shown that this is the case even for relatively young children.

These results contradict any linguistic theory of individual sentence memory, for they show that subjects do not store individual sentences at all. Instead, the individual sentences are combined both with each other and with subjects' generalized knowledge of the world to form an abstract representation of the general idea underlying the passage. As important as this work is, however, it should be clear that this is only a first step. We are still lacking an adequate formal description of the nature of these abstract representations. In the words of Bransford and Franks (1971), "a very important problem . . . concerns the question of what is learned in the above situations. How can one characterize the nature of the semantic ideas that are acquired?" (p. 349).

A Test of Two Coding Strategies

There are two different types of coding strategies which could account for subjects' inability to distinguish between information which had actually been presented and information which they, themselves deduced. The first was suggested by Quillian (1969) who noted that it would be efficient for subjects to delete from memory any information which was redundant in the sense that it could be deduced from other information stored in memory. This would be an efficient coding strategy for it would allow subjects to reduce their memory load without any corresponding loss of information; they could deduce any of the deleted material whenever they needed it. Using this coding strategy, however, subjects would not be able to determine if a particular piece of redundant information had actually been presented, for the redundant information would not be present in memory in either case.

A second alternative is that subjects make inferences while studying and that these inferences are stored along with the information which was actually presented. Since this would imply that redundant information is stored regardless of whether it was actually presented or not, such a theory could also account for subjects' inability to determine whether a particular piece of redundant information had actually been presented.

A series of experiments have been performed in our laboratory which were designed to determine whether or not subjects deduce and store inferences while studying a passage of text. The experimental paradigm employed in all of these experiments consisted of giving subjects one or more paragraphs to study. Each paragraph described a linear ordering of four terms. The terms were chosen and arranged in such a way that no "natural" ordering of the terms was apparent. In all cases, the relation was described by using the linguistically unmarked form of comparative adjective. Such an ordering (which will be characterized as $A > B > C > D$) can be described in terms of six pairwise relationships. Three of these pairs ($A > B$, $B > C$, and $C > D$) describe the relations between adjacent elements in the ordering and will be referred to as adjacent pairs. These adjacent pairs are necessary to the establishment of the ordering. The remaining three pairs ($A > C$, $B > D$, and $A > D$) describe the relations between nonadjacent elements in the ordering and will be referred to as remote pairs. Since the relations employed were transitive, these remote pairs could be deduced from some subset of the

adjacent pairs. Each paragraph presented only the three adjacent pairs necessary to establish a single ordering. The order of presentation of these pairs was varied across paragraphs. A sample paragraph is given below.

Sample Paragraph Describing a 4-term Linear Ordering

> In a small forest just south of nowhere, some animals were battling for dominion over the land. It boiled down to a battle of wits, so intelligence was the crucial factor. The bear was smarter than the hawk. The hawk was smarter than the wolf. The wolf was smarter than the deer. In the end the battle was decided and tranquility returned to the area.

A set of 12 test sentences (6 true and 6 false) was used to assess subjects' knowledge of the information contained in each of the paragraphs. The 6 true sentences consisted of a statement of the 6 pairwise relations comprising the ordering. For each true sentence (e.g., A > B?), there was a corresponding false sentence consisting of the same two terms in reverse order (e.g., B > A?). Test sentences all employed the linguistically unmarked form of comparative adjective.

The results were surprising in that proportion correct was higher and reaction time shorter on the remote pairs which had to be deduced than on the adjacent pairs which were actually presented. This apparent superiority of productive over reproductive memory is a very robust effect which has been replicated several times (Potts, 1972; Scholz & Potts, 1974; Potts, 1974a; Potts, 1974b). It should be clear that this result is completely inconsistent with the notion that the remote pairs are not stored, but are deduced from the adjacent pairs at the time of test. If this were the case, then proportion correct on a remote pair would have to be lower than proportion correct on any of the adjacent pairs necessary to deduce it, and hence overall proportion correct on the remote pairs would have to be lower than overall proportion correct on the adjacent pairs. Similarly, since deducing the remote pairs is known to take time (Huttenlocher, 1968; Clark, 1969), reaction time to these remote pairs would have to be longer than reaction time to the adjacent pairs.

Thus, the observed superiority on the remote pairs would appear to indicate that, at least under certain circumstances, subjects do successfully deduce and store inferences while studying textual material. In view of educational psychologists' intense interest in the processes involved in "going beyond the information given," one is led to question what it was about the present situation that led subjects to perform so well on the deducible information. This becomes an espeically interesting question when it is realized that some experimenters have found the exact opposite effect using a paradigm which, on the surface appears to be quite comparable to the one employed above.

A Conflicting Result

The studies in question were performed by Frase (1969, 1968, 1971). In these experiments, subjects learned a series of paragraphs describing a 5-term set inclusion relationship ($A \subset B \subset C \subset D \subset E$). A sample paragraph is given below. After studying the paragraphs, subjects were asked to recall all the information they could remember about the paragraphs. Subjects were instructed to write down not only information which was actually presented, but also any information that could be deduced from the presented material. Following this recall test, subjects were given a true-false recognition test similar to the one employed in our linear ordering experiments.

Sample Paragraph Describing a 5-Term
Set Inclusion Relation (from Frase, 1969)

> The Fundalas are outcasts from other tribes in Central Ugala. It is the custom in this country to get rid of certain types of people. The outcasts of Central Ugala are hill people. There are about fifteen different tribes in this area. The hill people of Central Ugala are farmers. The upper highlands provide excellent soil for farming. The farmers of this country are peace loving people, which is reflected in their art work.

Frase found that performance on the deducible remote pairs was poor relative to performance on the adjacent pairs. On the recall test, this difference was very large and highly significant. On the recognition test this difference, though in the same direction, was small and not significant. In view of this poor performance on the remote pairs, Frase concluded that subjects did not generally deduce and store inferences while studying. This result and conclusion is, of course, diametrically opposed to the result and conclusion drawn on the basis of our linear ordering results.

Experiment 1 was designed to determine the reason for the apparent discrepancy between our original linear ordering results and the results obtained by Frase.

Experiment 1

There are three immediately apparent differences between our experiments and those of Frase. (1) The most obvious is the fact that our experiments used linear orderings (i.e. "A is taller than B") and Frase's experiments used set inclusion relations (i.e. "All A are B"). Though the logical characteristics of the two types of relation are essentially the same, it is quite reasonable that subjects may not encode these two relations in the same way. (2) Secondly, examination of the two sample paragraphs reveals that in Frase's experiment extraneous information

(e.g. "It is the custom in this country to get rid of certain types of people.") was inserted between the four critical sentences describing the adjacent relations. In our experiments no such extraneous material intervened between the critical sentences. It is possible that the presence of extraneous information in Frase's experiment may have hindered subject's ability to make inferences. (3) Finally, our linear ordering experiments all employed true-false recognition tests. Frase found that though recognition performance on remote pairs was worse than recognition performance on adjacent pairs, the difference was small and non-significant. The inferiority on the remote pairs was most striking on the recall test. It may be that the type of test used is the critical factor determining subjects ability to demonstrate their knowledge of inferred material. Experiment 1 was designed to evaluate the importance of each of these factors.

Method

Subjects. Subjects were 64 Dartmouth College undergraduates who participated to fulfill a course requirement. Each subject participated in one 40 minute session.

Materials. Each subject studied a total of six paragraphs. Three of these paragraphs described a 5-term linear ordering, three described a 5-term set inclusion relation. In order to eliminate primacy and recency effects, two of the paragraphs (one linear ordering and one set inclusion) were used as fillers, one at the beginning and one at the end of the task.

The linear ordering paragraphs were modeled after those used in our original experiments. The set inclusion relations were modeled after those employed by Frase with one possibly important difference. In Frase's experiments, the universal quantifier ("all") was not always stated explicitly. Hence, some sentences read "All the outcasts of Central Ugala are hill people," while others omitted the quantifier and read simply "The outcasts of Central Ugala are hill people." In the present paragraphs, the universal quantifier was always stated explicitly.

The relationships were established by presenting the four adjacent pairs in chained order (A > B, B > C, C > D, D > E). In addition to these five essential sentences, the paragraphs also included some extraneous material. For half the paragraphs, this extraneous material was interspersed among the five essential sentences; for the other half of the paragraphs the extraneous material was inserted in a block following the five essential sentences. The specific linear ordering and set inclusion paragraphs used for each of these conditions was counterbalanced across subjects as was the order of presentation of the four paragraph types.

Procedure. Subjects were run in large groups. Each subject was given a test booklet containing the paragraphs and test materials. Subjects were given 75 sec. to study each paragraph. They were allowed to take notes while studying each paragraph, but it was made clear to them that they would not be allowed to use the notes while they were being tested. Subjects studied all six paragraphs before they were tested on any of the information.

After studying all six paragraphs, subjects were given a recall test on the information. Their test booklet contained six pages which were blank except for a title at the top of each page indicating which paragraph subjects were to recall. The first page contained the title of the first paragraph. Subjects were given 5 minutes to recall all the information they could remember about this paragraph. After this five minute period, subjects went on to the second page and recalled all the information they could remember about the second paragraph. This procedure was continued until subjects had recalled all six paragraphs. Subjects were instructed to respond with simple declarative sentences and to list not only the information that was actually presented but also any information that could be deduced from the presented information. These recall instructions were comparable to those employed by Frase (1969). Subjects in the present experiment were also given a concrete example to clarify the instructions.

Following the recall test, subjects were given a true-false recognition test of their memory for the information contained in each of the paragraphs. The test materials for each paragraph consisted of a set of 20 sentences (10 true and 10 false) each describing a relationship between a pair of terms in the paragraph. The 10 true sentences consisted of a statement of the four adjacent and six remote pairs comprising the 5-term relationship. Corresponding to each true sentence (e.g. A > B?) was a false sentence employing the same two terms in reverse order (e.g. B > A?). Subjects responded by writing "True" or "False" next to each of the 20 sentences for the first paragraph. They then turned a page and responded to the sentences for the second paragraph. This procedure was repeated until subjects had responded to the sentences for all six paragraphs. The sentences for each paragraph were listed on a separate sheet of paper. A computer was used to generate a different random order of presentation of the sentences for each paragraph and for each subject.

Results

The results did not differ in any way as a function of whether the extraneous information was interspersed among the critical sentences or included in a block at the end of the paragraph. Hence, all the results described below are averaged over that variable. Mean proportions correct on the recall and recognition tests are given in Table 1. Statistical analyses were performed using a sign test for matched samples. This is the most appropriate statistical measure given the binomial nature of the data.

Insert Table 1 about here

As can be seen from Table 1, the recognition results replicate our previous linear ordering results in that proportion correct was higher on the remote pairs than on the adjacent pairs. This difference was significant for both linear orderings and set inclusion relations ($z = 2.55$, $p = .01$ and $z = 2.30$, $p = .02$, respectively).

The recall results replicate those obtained by Frase in that recall of the remote pairs was very poor relative to recall of the adjacent pairs. This difference was significant for both linear orderings and set inclusion relations ($z = 3.43$, $p < .001$ and $z = 4.50$, $p < .001$, respectively).

Discussion

Recognition results. The present finding that recognition of remote pairs is more accurate than recognition of adjacent pairs for set inclusion relations as well as for linear orderings strongly suggests that for both types of relation subjects deduce and store inferences while studying textual material. While this result does not actually contradict Frase's recognition results which found no significant difference between remote and adjacent pairs, a replication of the present results would be desireable. This is especially true in view of the fact that several features of the present experiment may have served to facilitate performance on the remote pairs.

First, it should be noted that in the present experiment subjects learned both linear orderings and set inclusion relations. It is possible that the presence of the linear ordering paragraphs affected the strategies which subjects used to encode the set inclusion material. Second, subjects in the present experiment were allowed to take notes while studying the material. This may have served to facilitate the drawing of inferences. Finally, the testing procedure employed in the present experiment was modeled after the one used by Frase. Specifically, subjects were given both a recall and a recognition test in that order. Since recall of the adjacent pairs was better than recall of the remote pairs, it is quite unlikely that the prior recall could account for the accurate recognition of remote pairs. It is reasonable, however, that the prior recall may have affected subsequent recognition in some way. Hence, it would be useful to obtain a measure of recognition performance that was not confounded in this way.

In addition to its major objective described below, Experiment 2 addresses all of the above criticisms. No linear ordering paragraphs were used, and subjects were not allowed to take notes while studying. Each subject was given only one test on the critical information. This test was either a recall or true-false recognition test.

Recall results. Though recognition of inferences (remote pairs) was quite good, recall of this information was very poor for both linear ordering and set inclusion relations. This result is hard to interpret in the case of the linear ordering relations, for, despite our instructions to the contrary, many subjects recalled this information merely by listing the ordering of the five terms (i.e. A > B > C > D > E). Though subjects responding in this way were not given credit for accurate recall of the remote pairs, one is inclined to suspect that they could have provided these pairs if they so desired. It is interesting to note that almost no subjects showed this kind of protocol when recalling set inclusion relations.

Even though subjects did not recall set inclusion relations by merely listing the terms in order, it is still possible that the failure to recall inferences relects the operation of a simple response strategy. Once subjects have output several adjacent pairs (e.g. A > B, B > C), there may be a tendency to repress the output of information that could be deduced from these pairs (e.g. A > C) even though they may be well aware of the fact that these are legitimate deductions from the presented information. The fact that subjects were explicitly instructed to list all deducible information (and were given a concrete example to illustrate) may lead one to question the likelihood of this explanation, but it certainly does not eliminate the possibility entirely.

To examine the viability of this suggestion, a revised method of scoring the recall protocols was employed. According to this method, subjects were given credit for recalling a remote pair (e.g. A > C) whenever they recalled some set of other pairs sufficient to deduce that remote pair (e.g. A > B, B > C). The assumption underlying this method of scoring is that subjects could have output the deduction but instead adapted a response strategy of suppressing that output. For linear orderings, this revised method of scoring raised proportion correct on the remote pairs from .427 to .754. Since this revised score on the remote pairs did not differ significantly from proportion correct on the adjacent pairs (z = .53, p = .60), this result is difficult to interpret. The high proportion correct on the adjacent pairs would necessarily lead to a very high proportion correct on the remote pairs using this method of scoring. The revised recall score is more informative for the set inclusion relations. Under this method of scoring, recall of remote set inclusion relations is raised from .314 to .396. In spite of the substantial improvement, recall of the remote pairs is still significantly lower than recall of the adjacent pairs ($\underline{z}$ = 3.71, p < .001). Hence, at least for the set inclusion relations it is very unlikely that the poor recall of the remote pairs can be accounted for entirely on the basis of a response bias against outputting pairs that are deducible from previously recalled items.

<u>Summary</u>. The results of Experiment 1 indicate the following:

(1) It makes virtually no difference (for either linear orderings or set inclusion relations) whether extraneous information is inserted between the critical sentences of a paragraph or is included in a block at the end of the paragraph. This is clearly not the factor responsible for the apparent discrepancy between our original results and those of Frase.

(2) Though overall performance on set inclusion relations is much poorer than performance on linear orderings, both types of relation act similarly with regard to the comparison between productive and reproductive memory (i.e. remote vs. adjacent pairs).

(3) Type of test (recall vs. true-false recognition) appears to be the most critical factor affecting relative performance on remote and adjacent pairs. Recognition memory for remote pairs was significantly better than recognition memory for adjacent pairs; recall of remote pairs was much worse than recall of adjacent pairs. This was

true for both linear orderings and set inclusion relations. Experiment 2 will examine one possible explanation for the very poor recall of the deducible remote pairs. In addition, it will address the points discussed previously as possible explanation of the very good recognition performance on these same remote pairs.

Experiment 2

Experiment 1 indicated that recognition of deducible remote pairs was quite good while recall of those same pairs was very bad. Several memory theories (e.g. Kintsch, 1972) have assumed that the major difference between recall and recognition tests is that the latter minimizes or eliminates the need for subjects to search through memory in an attempt to retrieve the desired information. It is assumed that the recognition test item provides sufficient information to allow subjects to access the appropriate part of memory. Hence, according to these theories, an item which can be correctly recognized but cannot be recalled is available (i.e. stored in memory) but not accessible (i.e. not able to be retrieved). This view would lead one to argue that though a subject has stored information about the remote pairs, he is unable to retrieve this information when tested for recall. Why should retrieval of this information be so difficult?

One possible explanation is that for some reason subjects may not have been expecting a recall test and therefore may not have coded their inferences in such a way as to facilitate easy retrieval. Experiment 2 examined this possibility by manipulating subjects' expectancy regarding the type of test they would be given. In addition, Experiment 2 included a delayed test administered one week after the initial session to determine the long term effects of this variable.

Method

<u>Subjects</u>. Subjects were 111 Dartmouth College undergraduates who participated to fulfill a course requirement. Each participated in two sessions, an initial 50 min. learning and test session, and a brief delayed test session held one week later.

<u>Material</u>. Three critical paragraphs were employed. Each described a single 5-term set inclusion relation similar to the ones employed in Experiment 1. Except for brief introductory and concluding sentences, no extraneous material was included. Both the recall and the recognition tests had the same characteristics as the ones employed in Experiment I.

In addition to the three critical paragraphs, six practice paragraphs were used. These six paragraphs described some artificial information about the effects of pot, skiing as a sport, technological innovations for students, the effects of music on plants, the politics of the republic of Dax, and new sources of power, respectively. None of the information in these practice paragraphs bore any resemblance to either linear orderings or set inclusion relations. Recognition tests for each of these paragraphs consisted of a set of eight (four true and four false) statements about the information described in the paragraph.

Two of the true sentences described information which was actually presented in the paragraph, two described information which could be deduced from information described in the paragraph.

Procedure. Subjects were run in large groups. Each subject has his own test booklet which included all learning and test materials. The initial learning and test session was broken into three parts. During each of these three parts, subjects studied and were tested on a different set of three paragraphs. The first two sets consisted of practice paragraphs; the last set contained the critical paragraphs.

After reading the initial instructions, subjects turned a page and began studying the first practice paragraph. After studying the first paragraph for 60 sec., the experimenter called time and subjects began studying the second paragraph. Study time for each paragraph was reduced to 60 sec. in this experiment because the critical paragraphs were shorter due to the deletion of the extraneous material. Subjects were not allowed to take notes while studying the paragraphs. After studying each of the three practice paragraphs in the first set, subjects were given either a recall or a recognition test on the information in that set of paragraphs. Subjects were allowed a total of 6 min. to complete the test on the three paragraphs.

Following the test on the first set of paragraphs, the whole procedure was repeated for the second set of three practice paragraphs. Subjects who received a recall test on the first set of paragraphs also received a recall test on the second set of paragraphs; subjects who received a recognition test on the first set also received a recognition test on the second set of paragraphs.

Following the test on the second set of practice paragraphs, subjects went on to study the set of three critical paragraphs. The study procedure for these paragraphs was identical to that followed for the practice paragraphs. After studying the paragraph however, subjects were given an additional set of instructions describing the kind of test (recall or recognition) they would receive on the critical set of paragraphs. This constituted the critical manipulation in the present experiment, and the four conditions are described in the "Design" section. Subjects were given as long as they needed to respond on this final test.

Following this test on the last set of paragraphs, subjects were dismissed and told to return for a second session one week later. When subjects returned for the second session, they were given a second test on the critical paragraphs. The type of this delayed test (recall or recognition) was in all cases the same as the type of the initial, immediate test on the critical paragraphs. Subjects were not allowed to study the material again prior to this delayed test.

Design. Subjects were divided randomly into four groups which differed according to the type of test subjects were led to expect and the types of test they actually received. The "Recognition - Recognition" group (n = 27) received a recognition test on all trials. The "Recall-Recall" group (n = 27) received a recall test on all trials. Hence, for both of these groups the tests on the first two sets of practice paragraphs were the same as the tests on the final, critical set of paragraphs. It

follows that with regard to the critical paragraphs, these subjects expected the kind of test they actually received.

The "Recognition-Recall" group (n = 28) received a recognition test on the first two sets of practice paragraphs, but a recall test on the immediate and delayed tests on the critical set of paragraphs. The "Recall-Recognition" group (n = 29) received a recall test on the first two sets of practice paragraphs but a recognition test on the immediate and delayed tests on the critical set of paragraphs. Hence, with regard to the critical paragraph, subjects in these two groups were expecting a different type of test than the one they actually received.

Results

The immediate and delayed recall results for the "Recall-Recall" and "Recognition-Recall" groups are presented in Table 2a. The immediate and delayed recognition results for the "Recognition-Recognition" and "Recall-Recognition" groups are presented in Table 2b. Within subject statistical analyses were again performed using two-tailed sign tests for matched samples. Between subject comparisons were performed using Mann-Whitney U-tests. Since the number of subjects in each group was greater than 20, the two-tailed significance of the $\underline{U}$ statistic was evaluated using the normal approximation or $\underline{z}$ score.

Insert Table 2a & b about here

<u>Recall</u>. Averaged over groups, the immediate recall results replicate the results obtained in Experiment 1. Recall of the remote pairs was substantially worse than recall of the adjacent pairs ($\underline{z} = 5.36$, $p < .01$). Recall of remote and adjacent pairs did not differ on the delayed test ($\underline{z} = .85$, $\underline{p} = .4$), probably due to the fact that overall recall in this condition was so low.

On the immediate test, recall of the remote pairs was better when subjects were expecting a recognition test ("Recognition-Recall") than when they were expecting a recall test ("Recall-Recall"). Immediate recall of the adjacent pairs showed the opposite effect, being slightly better when subjects were expecting a recall test. This interaction approached, but did not reach significance ($U = 275.5$, $z = 1.73$, $p = .08$). The main effect of expectation did not approach significance ($U = 359.0$, $z = .32$, $p = .74$).

On the delayed test, recall of both remote and adjacent pairs was better when subjects were expecting a recognition test than when they were expecting a recall test, but this main effect of expectation was not significant ($U = 297$, $z = 1.36$, $p = .17$). The interaction also failed to reach significance for the delayed test ($U = 299$, $z = 1.33$, $p = .18$).

Recognition. Averaged over groups, the recognition results also replicate the results obtained in Experiment 1. Proportion correct on the remote pairs was higher than proportion correct on the adjacent pairs. This difference was not significant for the immediate test but was significant for the delayed test ($z = 1.37$, $p = .27$, and $z = 3.00$, $p < .01$, respectively).

The nature of this high recognition performance on the remote pairs is clarified by examining proportion correct as a function of the truth value of the test sentence. This data is presented in Table 3. As can be seen, proportion correct on true sentences did not differ for remote and adjacent pairs. The superiority on the remote pairs was entirely the result of performance on the false sentences. In other words, the tendency to say "True" to a true sentence did not differ for remote and adjacent pairs. However, subjects were much more likely to say "False" to a remote false sentence than they were to say "False" to an adjacent false sentence. This interaction between remote versus adjacent pairs and truth value of the test sentence was significant for both the immediate and delayed tests ($z = 3.04$, $p < .001$, and $z = 2.33$, $p = .02$, respectively).

Insert Table 3 about here

Proportion correct on both remote and adjacent pairs was better when subjects were expecting a recognition test ("Recognition-Recognition") than when they were expecting a recall test ("Recall-Recognition"). This main effect of expectancy was significant for the immediate test, but not for the delayed test ($U = 267.0$, $z = 2.05$, $p = .04$ and $U = 302.0$, $z = 1.47$, $p = .14$ respectively). The interaction between type of pair and expectancy did not approach significance for either the immediate or delayed test ($U = 352.0$, $z = .65$, $p = .52$ and $U = 355.0$, $z = .599$, $p = .54$ respectively).

Discussion

Effects of expectancy. Numerous studies have been performed examining differences in coding strategies when subjects are expecting recall or recognition tests. However, no one has looked explicitly at the effects of this variable on subject's ability to draw inferences from text. The major purpose of the present experiment was to determine whether the very poor recall of remote pairs observed by Frase and in Experiment 1 of the present series could be explained by arguing that subjects in those experiments were expecting a recognition test and hence did not store inferences in such a way as to facilitate retrieval of that information (the high recognition scores on these remote pairs indicated that this information was stored in some form). Hence, it was hypothesized that this poor recall of remote pairs might be accentuated for the "Recognition-Recall" group which was given a concrete reason to expect a recognition test, and reduced or eliminated for the "Recall-Recall" group which was led to expect a recall test.

This hypothesis was disconfirmed in that both recall and recognition of remote pairs was better when subjects were expecting a recognition test than when they were expecting a recall test. This effect was small, however, and was not significant in the case of the recall data. For the immediate recall test the interaction between type of pair and expectancy did approach significance, however. This reflected the fact that performance on adjacent pairs was better when subjects expected a recall test, while performance on remote pairs was better when subjects expected a recognition test. This would lead one to suspect that perhaps subjects who expected a recall test merely memorized the information that was actually presented while subjects who expected a recognition test had a greater tendency to go beyond the information given.

Because of the marginal significance of these effects, no firm conclusions regarding the role of expectancy can be drawn from this experiment. If further research shows these effects to be reliable, however, the results could prove to be of practical importance in determining those factors which facilitate the process of drawing inferences from a passage of text.

Since immediate recall of remote pairs was worse than recall of adjacent pairs regardless of the type of test subjects were expecting, we are still lacking an explanation for the poor recall of this information.

Recognition Performance. The recognition results of the present experiment replicated the results of Experiment 1 in that proportion correct on the remote pairs was higher than proportion correct on the adjacent pairs. This was the case even though subjects in the present experiment were not allowed to take notes. It was suggested that the presence of the linear ordering paragraphs in Experiment 1 may have affected the strategies which subjects used to encode the set inclusion material. Since no linear ordering paragraphs were used in Experiment 2 however, this could not account for the high recognition performance on the remote pairs in this experiment.

Though for both linear orderings and set inclusion relations proportion correct on remote pairs has been higher than proportion correct on adjacent pairs, it would be inaccurate to conclude from this that subjects process these two types of relations in a similar manner. The fact that overall performance on linear orderings is so much higher than performance on set inclusion relations clearly indicates a substantial difference in the way the two types of relations are processed. Subjects performance on adjacent and remote pairs as a function of truth value of the test sentence provides a clue to the nature of this difference.

Examination of Table 3 reveals that, for true sentences, performance on remote and adjacent pairs does not differ. The superiority on the remote pairs is due entirely to performance on the false sentences. This is not the case with linear orderings, where we typically find that performance on the remote pairs is superior to performance on the adjacent pairs regardless of the truth value of the test sentence.

Griggs (1974) has reported a series of experiments demonstrating similar effects. Grigg's results differed from the ones obtained in the present experiment only in that, for true sentences describing set inclusion relations, Griggs found proportion correct on remote pairs to be actually worse than proportion correct on adjacent pairs. Griggs argued that this profile reflected not memory errors but two different logical errors in dealing with set inclusion relations. (1) Subjects tended to erroneously assume that the relations were symmetrical (i.e., if told that "all A are B," they tended to respond true to the sentence "all B are A"). This accounts for the very low proportion correct on false statements of adjacent relations. This kind of "invalid conversion" (Johnson, 1972) error has also been observed by several other researchers studying syllogistic reasoning (e.g. Johnson, 1972; Ceraso & Provitera, 1971; Chapman & Chapman, 1959; Woodworth & Sells, 1935). (2) Subjects failed to assume transitivity (i.e., if told that "all A are B" and "all B are C," they tended to respond false to any sentence, true or false, employing the pair of terms A and C). This accounts for the fact that as inferential distance increased, performance on true sentences decreased while performance on false sentences increased.

Experiment 3

If linear orderings and set inclusion relations differ only in that subjects tend to make logical errors when responding to set inclusion relations, then it is reasonable to suspect that if subjects are given a chance to correct their logical errors, proportions correct for the set inclusion relations should become very good and the reaction times obtained using this relations should match those obtained with linear orderings. Experiment 3 examined reaction time profiles obtained with linear orderings and set inclusion relations.

These reaction time results are very important theoretically. The fact that proportion correct on remote pairs is higher than proportion correct on adjacent pairs strongly suggests that subjects are deducing and storing these inferences while reading the passage. This contradicts several models (e.g. Quillian, 1969) which argue that, in the interests of cognitive economy, deducible information is inferred at the time of test rather than being stored directly. Other models, however, have argued that subjects do, indeed, store inferred information (e.g. Rumelhart, Lindsay, & Norman, 1972; Anderson & Bower, 1973). For example, Anderson and Bower (p. 407) have proposed that in addition to learning the adjacent pairs comprising a linear ordering, subjects may also learn some remote relations. They maintain Ebbinghaus' (1885) contention that the more remote an association, the weaker it is. Such a model can account for the fact that proportion correct is higher on the remote pairs than on the adjacent pairs. In order to be correct on an adjacent pair, subjects would have to remember that pair. Subjects could be correct on a remote pair either by remembering that pair or by remembering some subset of other pairs which is sufficient to deduce it. However, since subjects responding correctly to a remote test sentence do so by either remembering the weak remote association or by deducing that relation from other pairs, such a model would have to predict a long reaction time

on these remote pairs. Our previous reaction time results have clearly contradicted this prediction for linear orderings (Potts, 1974b). We have found that reaction time to the remote pairs of a linear ordering is consistently shorter than reaction times to the adjacent pairs. Experiment 3 will determine whether this is also the case for set inclusion relations.

Method

Subjects. Subjects were 16 Dartmouth College undergraduates who participated to fulfill a course requirement. Each subject participated in a single 30 min. session.

Materials. Each subject studied and responded to a set of eight critical paragraphs. Each paragraph described a single 4-term linear ordering or set inclusion relation. In order to make the linear ordering and set inclusion relations as similar as possible, nonsense syllables were used for the terms of each relation. Hence, for example, a sentence describing a linear ordering relation would read

DAX are bigger than MEP

while a sentence describing a set inclusion relation would read

All DAX are MEP

The three adjacent pairs describing each relation were presented in the chained order AB, BC, CD.

The test on each relation consisted of a statement of the 12 possible pairwise relations comprising a four term ordering. The 6 true sentences consisted of a statement of the three adjacent and three remote pairs of the ordering. Corresponding to each true sentence (e.g. A > B?) there was a false sentence which described the same two terms in reverse order (e.g. B > A?)

Design. Subjects were divided randomly into two groups. Subjects in one group (n = 8) studied and responded to nine paragraphs describing 9 different linear ordering relations (e.g. "DAX are bigger than MEP"). Subjects in the second group (n = 8) studied and responded to nine paragraphs describing set inclusion relations (e.g. "All DAX are MEP"). The four nonsense syllables used in each relation were the same for both groups. Hence, the paragraphs for the two groups differed only in the relation employed.

Procedure. Subjects were run one at a time. After a brief practice paragraph, the experimenter gave the subject the first critical paragraph to study. Subjects were allowed to study the paragraph for as long as they liked, and were allowed to take notes if they so desired. When the subject indicated he was ready, the experimenter took away the paragraph and notes, and the subject began responding to the set of 12 test sentences for that paragraph.

The sentences were presented one at a time using the paper advance mechanism from a high speed line printer. The subject responded to each sentence by pressing one of two buttons on the response box in front of him. For half the subjects in each condition, the left and right buttons were labeled "True" and "False" respectively. For the other half of the subjects the labeling was reversed. Reaction times were measured using a Lafayette millisecond timer which was started simultaneously with the presentation of the sentence and stopped automatically when the subject responded. Subjects were instructed to respond as quickly as possible to each sentence, but to be very careful not to respond so quickly as to cause them to make errors.

After responding to all 12 sentences from the first paragraph, subjects were informed of the number of errors they had made. It was hoped that this would serve to enable them to correct any inappropriate responding strategies. After receiving feedback, subjects were given the second paragraph to study. This procedure was repeated until subjects had studied and responded to all eight critical paragraphs. The sentences were listed on a computer printout and the order of presentation of the 12 test sentences was randomly permuted for both paragraphs and sentences.

Results

Overall proportion correct was extremely high. Subjects receiving linear ordering paragraphs had an overall proportion correct of .954; subjects receiving set inclusion paragraphs had an overall proportion correct of .958. Table 4 presents the mean reaction times in these two conditions for true and false sentences. As can be seen, reaction time is shorter on the remote pairs for both linear orderings and set inclusion relations. This is the case for both true and false sentences. The significance of these effects was tested using a 3 way analysis of variance. Type of pair (adjacent vs. remote) and truth value of the test sentence were within subject variables, type of relation (linear ordering versus set inclusion) was a between subject variable. Since each score in the analysis represented an average over several responses, the resulting reaction time scores should not deviate seriously from normality.

Insert Table 4 about here

Reaction time to the remote pairs was significantly shorter than reaction time to the adjacent pairs, $F(1, 14) = 30.11$, $p < .001$. Reaction time to true sentences was significantly shorter than reaction time to false sentences, $F(1, 14) = 21.11$, $p < .001$. No other main effects or interactions approached significance ($p > .10$ in all cases). Of major importance is the fact that no effects involving the comparison between linear orderings and set inclusion relations were significant.

Experiment 4

One could criticize the previous experiment on the grounds that the use of nonsense syllables served to make the materials more artificial than the materials used in previous experiments. It could be argued that this may have affected subject's strategies. Experiment 4 examined reaction times to answer questions about set inclusion relations which were more similar to the ones employed in Experiments 1 & 2.

Method

Subjects were 16 Dartmouth College undergraduates who participated to fulfill a course requirement. Each participated in two 30 min. sessions. Four paragraphs were employed. Each described a single four term set inclusion relation by presenting the three adjacent pairs in chained order. The relations had the same characteristics as those employed in Experiments 1 and 2 (except of course that the present relations consisted of only four terms). The test materials had the same characteristics as those employed in Experiment 3.

After a brief practice paragraph, subjects were given four trials on the first paragraph followed by four trials on the second paragraph. That completed the first session. One week later subjects returned and were given a retest on the first two paragraphs. They were then given four trials on the third paragraph followed by four trials on the fourth paragraph. Feedback was again given after each paragraph.

Except for these differences, the procedures employed were identical to those employed in Experiment 3.

Results

No interesting effects of trials was noted (except for an overall increase in the speed of responding over trials), so results were averaged over trials and paragraphs. Overall proportion correct was .911. Table 4 presents mean reaction times to adjacent and remote pairs as a function of type of pair (adjacent vs. remote) and truth value of the test sentence. As can be seen, reaction time to the remote pairs was again shorter than reaction times to the adjacent pairs. This was the case for both true and false sentences. The short reaction time to the adjacent pairs was demonstrated by 14 or the 16 subjects and was thus highly significant by a two tailed sign test ($p < .01$). It should be noted that overall reaction times in the present experiment were substantially longer than in Experiment 3. This was probably due to the fact that the sentences in the present experiment were much longer and more complicated than the sentences in the previous experiment (e.g. "All *DAX* are *MEP*" vs. "All the red-eyed birds of Brazil are water birds").

Discussion

Experiment 3 indicates that when subjects are given the chance to correct their logical errors, proportion correct on set inclusion relations becomes very good and the resulting reaction time profiles for set inclusion relations do not differ from the profile obtained for a linear ordering. In both cases reaction time to remote pairs is shorter than reaction time to adjacent pairs. Experiment 4 demonstrated the same effect using materials that were more comparable to those used in Experiments 1 and 2.

This result leads one to the interesting hypothesis that perhaps set inclusion and linear ordering relations are both represented in the same form (perhaps an ordering of the terms) and that the only difference between them is the tendency for subjects to employ inappropriate response strategies when answering set inclusion questions. We are currently engaged in a series of experiments designed to test this possibility.

The fact that reaction time to the remote pairs of a set inclusion relation was shorter than reaction time to the adjacent pairs confirms our previous results using linear orderings, and would appear to provide further evidence against an associative model of the type proposed by Anderson and Bower (1973). However, our linear ordering results indicate that subjects apparently learn the end terms of such an ordering and are able to use this information to make a fast response (c.f. Potts, 1972; Potts, 1974b for details). If this is also the case for set inclusion relations, then a definitive test of Anderson and Bower's model using set inclusion relations would require the use of an ordering of more than four terms so that we could examine distance effects among pairs which do not include either end term. We are currently examining this possibility.

Conclusions

The conclusions to be drawn from the present series of experiments can be summarized as follows:

1) The present experiments indicate that, using a true-false recognition test, proportion correct is higher and reaction time shorter on the remote pairs of an ordered relationship (i.e. productive memory) than on the adjacent pairs (i.e. reproductive memory) of such a relationship. This is the case for both linear orderings and set inclusion relationships. This surprisingly good productive memory strongly indicates that when studying text which describes a linear ordering or set inclusion relation, subjects make and store inferences as an integral part of the reading process.

2) The presence or absence of extraneous material between the adjacent pairs necessary to construct an ordered relationship appears to have no effect on subject's ability to make and store inferences.

3) Though productive memory was found to be better than reproductive memory when measured using a true-false recognition test, productive memory was found to be much worse than reproductive memory when measured using a recall test. A suggested explanation for this effect was that though subjects made inferences while studying the text, these inferences may not have been stored in such a way as to facilitate easy retrieval of the information. This could be expected to hinder recall of this information more than it would hinder recognition.

4) The above hypothesis could lead one to suspect that the poor recall of deducible information might be alleviated by making it clear to subjects that they would be tested using a free recall testing procedure. This was not the case. Recall of inferred information was slightly (though not significantly) better when subjects were expecting a recognition test than when they were expecting a recall test (which they actually received). Immediate recall of the adjacent pairs showed the opposite effect, being

somewhat better when subjects expected a recall test. This suggests that perhaps subjects try to memorize exactly what was presented when they expect a recall test but have a tendency to "go beyond the information given" when expecting a recognition test. Though these expectancy effects are interesting, they are unfortunately also quite small and of questionable reliability. Further research is necessary before any definitive statements about the role of expectancy can be made.

Extensions

The present experiments raise several questions regarding the way in which subjects process ordered relationships in particular and text in general.

1) The present experiments do not address themselves to the question of why productive memory is so good when learning ordered relationships. This is a very important question which deserves careful examination.

2) A related question deals with the exact nature of the differences between linear orderings and set inclusion relations. The present experiments lead one to speculate that perhaps these two relations are processed and stored in the same way but that subjects tend to make logical errors (Griggs, 1972) when responding to set inclusion questions. Additional research will be necessary before we are able to evaluate the viability of such a suggestion.

3) The present experiments suggest that leading subjects to expect a true-false recognition test (as opposed to a recall test) may improve their ability to recall deducible information. This effect could turn out to have considerable educational relevance with regard to the type of test which may be most effective in facilitating the process of "going beyond the text." Further research is essential, however, to demonstrate the reliability of this effect, its generality, and its size. Not until this research is performed will we be able to determine the practical significance of the finding.

REFERENCES

Anderson, J. R. & Bower, G. H. Human Associative Memory. Washington: Wiley, 1973.

Barclay, J. R. The role of comprehension in remembering sentences. Cognitive Psychology, 1973, 4, 229-254.

Bartlett, F. C. Remembering. Cambridge: Cambridge University Press, 1932.

Blumenthal, A. L. Observations with self-embedded sentences. Psychonomic Science, 1966, 6, 453-454.

Blumenthal, A. L. & Boakes, R. Prompted recall of sentences, a further study. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 674-676.

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory: A constructive versus interpretive approach. Cognitive Psychology, 1972, 3, 193-209.

Bransford, J. D. & Franks, J. J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Ceraso, J. & Provitera, A. Sources of error in syllogistic reasoning. Cognitive Psychology, 1971, 2, 400-410.

Chapman, L. J. & Chapman, J. P. Atmosphere effect re-examined. Journal of Experimental Psychology, 1959, 58, 220-226.

Chomsky, N. Syntactic Structures. The Hague: Mouton, 1957.

Chomsky, N. Aspects of the Theory of Syntax. Cambridge: M.I.T.Press, 1965.

Clark, H. H. Linguistic processes in deductive reasoning. Psychological Review, 1969, 76, 387-404.

Ebbinghaus, H. Memory: A Contribution to Experimental Psychology. Teacher's College, Columbia University, New York, 1885.

Frase, L. T. Some unpredicted effects of different questions upon learning from connected discourse. Journal of Educational Psychology, 1968, 59, 197-201.

Frase, L. T. Structural analysis of the knowledge that results from thinking about text. Journal of Educational Psychology, 1969, 60 (monograph supplement 6).

Frase, L. T. Influence of sentence order and amount of higher level text processing upon reproductive and productive memory. American Educational Research Journal, 1970, 7, 307-319.

Gough, P. B. Grammatical transformation and speed of understanding. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 107-111.

Griggs, R. A. Logical errors in comprehending set inclusion relations in meaningful text. Unpublished doctoral dissertation, Indiana University, 1974.

Huttenlocher, J. Constructing spatial images: A strategy in reasoning. Psychological Review, 1968, 75, 550-560.

Johnson, D. M. Systematic Introduction to the Psychology of Thinking. New York: Harper & Row, 1972, 229-271.

Kintsch, W. Models for free recall and recognition. In Norman, D. A. (Ed.), Models of Human Memory. New York: Academic Press, 1970, 331-373.

Martin, E. & Roberts, K. H. Grammatical factors in sentence retention. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 211-218.

Mehler, J. Some effects of grammatical transformation on the recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 346-351.

Mehler, J. & Miller, G. A. Retroactive interference in the recall of simple sentences. British Journal of Psychology, 1964, 55, 295-301.

Miller, G. A. Some psychological studies of grammar. American Psychologist, 1962, 17, 748-762.

Paris, S. G. & Carter, A. Y. Semantic and constructive aspects of sentence memory in children. Developmental Psychology, 1973, 9, 109-113.

Potts, G. R. Information processing strategies used in the encoding of linear orderings. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 727-740.

Potts, G. R. Incorporating quantitative information into a linear ordering. Memory and Cognition, 1974, 2, 533-538. (a)

Potts, G. R. Storing and retrieving information about ordered relationships. Journal of Experimental Psychology, 1974, 103, 431-439. (b)

Quillian, M. R. The teachable language comprehender: A simulation program and theory of language. Communications of the ACM, 1969, 12, 459-476.

Rohrman, W. L. The role of syntactic structure on the recall of English nominalizations. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 904-912.

Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), Organization of Memory. New York: Academic Press, 1972, 197-246.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442.

Savin, H. B. & Perchonock, E. Grammatical structure and the immediate recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 348-353.

Scholz, K. W. & Potts, G. R. Cognitive processing of linear orderings. Journal of Experimental Psychology, 1974, 102, 323-326.

Woodworth, R. S. & Sells, S. B. An atmosphere effect in formal syllogistic reasoning. Journal of Experimental Psychology, 1935, 18, 451-460.

Table 1

Mean Proportions Correct on Recall and

Recognition Tests in Experiment 1

	Recall Test		Recognition Test	
Type of Material	Adjacent Pairs	Remote Pairs	Adjacent Pairs	Remote Pairs
Linear Ordering	.771	.427	.919	.951
Set Inclusion	.492	.314	.675	.718

Table 2a

Recall Performance on Immediate and Delayed Tests as a Function of Type of Test Expected

	Type of Test Expected			
	Recall		Recognition	
	Adjacent Pairs	Remote Pairs	Adjacent Pairs	Remote Pairs
Immediate Test	.719	.345	.651	.440
Delayed Test	.351	.236	.413	.389

Table 2b

Recognition Performance on Immediate and Delayed Tests as a Function of Type of Test Expected

	Type of Test Expected			
	Recall		Recognition	
	Adjacent Pairs	Remote Pairs	Adjacent Pairs	Remote Pairs
Immediate Test	.748	.776	.830	.864
Delayed Test	.659	.717	.719	.786

Table 3

Recognition Performance on Immediate and Delayed Tests as a Function of Truth Value of the Test Sentence

	TRUE SENTENCES		FALSE SENTENCES	
	Adjacent Pairs	Remote Pairs	Adjacent Pairs	Remote Pairs
Immediate Test	.886	.872	.691	.767
Delayed Test	.741	.767	.636	.736

Table 4

Reaction Times on Linear Orderings and Set Inclusion Relations in Experiment 3 & 4

	Truth Value of Test Sentence			
	TRUE		FALSE	
	Adjacent Pairs	Remote Pairs	Adjacent Pairs	Remote Pairs
Linear Ordering	1.86	1.54	2.01	1.87
Set Inclusion (Experiment 3)	1.69	1.48	2.01	1.71
Set Inclusion (Experiment 4)	2.64	2.45	2.88	2.53